IMPLEMENTATION AND PERCEPTUAL EVALUATION OF A SIMULATION METHOD FOR COUPLED ROOMS IN HIGHER ORDER AMBISONICS

Giso Grimm,
Medical Physics
Universität Oldenburg and
Cluster of Excellence "Hearing4all",
Oldenburg, Germany
g.grimm@uni-oldenburg.de

Torben Wendt,
Acoustics and Medical Physics
Universität Oldenburg and
Cluster of Excellence "Hearing4all",
Oldenburg, Germany
torben.wendt@uni-oldenburg.de

Volker Hohmann,
Medical Physics
Universität Oldenburg and
Cluster of Excellence "Hearing4all",
Oldenburg, Germany
volker.hohmann@uni-oldenburg.de

Stephan D. Ewert,
Medical Physics
Universität Oldenburg and
Cluster of Excellence "Hearing4all",
Oldenburg, Germany
stephan.ewert@uni-oldenburg.de

ABSTRACT

A fast and perceptively plausible method for rendering acoustic scenarios with moving sources and moving listeners is presented. The method is principally suited for application in dynamic and interactive evaluation environments (e.g., for hearing aid development), psycho-physics with adaptively changing the spatial configuration, or simulation and computer games. The simulation distinguishes between the direct sound, sound reflected and diffracted by objects of limited size, diffuse sound surrounding the listener, e.g., diffuse background sounds and diffuse reverberation, and 'radiating holes' for simulation of coupled adjacent rooms. Instead of providing its own simulation of room reverberation, the proposed simulation method generates appropriate output signals for external room reverberation simulators (e.g., see contribution by Wendt et al.). The output of such room reverberation simulators is then taken either as diffuse surrounding sound if the listener position is within the simulated room, or as input into a 'radiating hole', if the listener is in an adjacent room.

Subjective evaluations are performed by comparing measured and synthesized transitions between coupled rooms.

1. INTRODUCTION

A large variety of tools for acoustic simulation of rooms and open spaces exists (e.g., [1], [2]). Most of these tools can simulate room acoustics at a very high accuracy. However, the required high complexity allows only to simulate static impulse responses or soundscapes. The time-varying simulation of acoustic spaces in real-time implies strong simplifications to typically used approaches of geometric acoustics (e.g., image sources [3]) and ray tracing ([4]) as well as combinations (e.g., [5]). Alternatively, prerendered acoustic scenes can be used (e. g. [6]). Dramatic simplifications are typical in real-time applications like computer games which seek immersive soundscapes at cost of realism. In this case often simple reverberation algorithms are used with settings predefined by the game designer. However, in opposite to static, physically accurate room simulation plausibility can, for specific applications, be more important than exact correspondence to the real world. Whereas a plausible approximation of room acoustics is often possible by the use of rectangular "shoebox" rooms for early reflection simulation in combination with a model for diffuse reverberation, the simulation of connected (coupled) rooms is more computationally demanding, nevertheless required for many auralization applications.

Acoustic simulations can be physically assessed by measures related to the simulated binaural room impulse response (e.g., decay time, early decay time, interaural cross correlation). Especially the simulation of existing environments can be compared with recorded impulse responses. Plausibility of acoustic simulations, however, is a more vague criterion. It may be measured by the perceptual component of auditory spatial awareness [7]: If a listener is able to identify a simulated environment purely by its acoustics, then a simulation may be seen as plausible.

In the present paper, the room coupling model of a toolbox for dynamic real-time simulation of acoustic spaces for hearing research and hearing aid evaluation (toolbox for acoustic scene creation and rendering, TASCAR, [8]) is described and perceptually evaluated. The toolbox aims to provide a physically accurate simulation of the direct sound and the first order reflections, together with a plausible reproduction of diffuse reverberation, diffuse environmental sounds and room coupling.

In Section 2, the simulation methods used by the toolbox are briefly described. Section 3 describes the evaluation methods used in this study and the measurement of real-world binaural impulse response that are used as a reference. The results (Section 4) are discussed and summarized.

2. SIMULATION METHODS

The proposed simulation method for dynamic transitions between coupled rooms is part of the TASCAR toolbox for acoustic scene creation and rendering [8]. The focus of this toolbox is the time-varying simulation of acoustics, i.e., all sound sources, receivers and reflecting objects can move dynamically, and the simulated acoustics is returned as a time signal, depending on the input signals and the time-varying spatial configuration of the simulated environment. In this framework, point sources follow a distance model with a $1/r$ sound pressure law, r being the distance between sound source and receiver. Additionally, air absorption is approximated by a simple first order low-pass filter model:

$$y_k = a_1 y_{k-1} + (1 - a_1) x_k \quad (1)$$

$$a_1 = e^{-\frac{r f_s}{c\alpha}}, \quad (2)$$

where c is the speed of sound. The empiric constant $\alpha = 7782$ was manually adjusted to provide sensible values for distances below 50 meters. The resulting absorption as a function of distance is given in Figure 1. This approach is very similar to that of [9] who used a FIR filter to model the frequency response at certain distances. However, in this approach the distance parameter r can be varied dynamically.

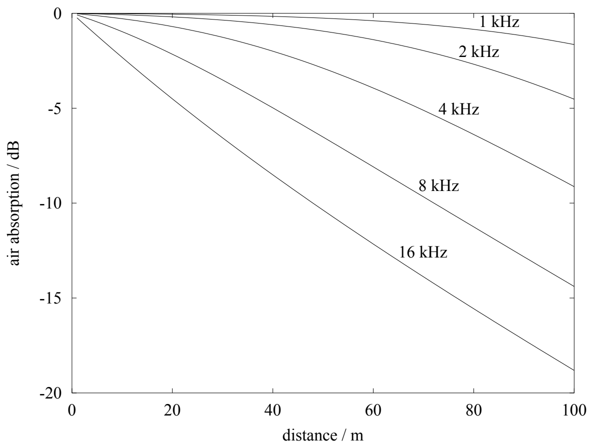

Figure 1: *Simulated air absorption as a function of distance. For distances below 50 m the simulated absorption in dB is roughly proportional to the distance. For larger distances, the low frequency absorption is too high.*

Early reflections are modeled using a image source model. In opposite to most commonly used models (e.g., [3]) which calculate impulse responses for a rectangular enclosure ("shoebox model"), reflections are simulated for each reflecting (rectangular, but arbitrarily oriented) surface. Diffraction is simulated by an empirically found attenuation g, which depends on the position of the image source $\mathbf{x}_{src}$, the receiver position $\mathbf{x}_{rec}$, and the point on the reflecting surface with the shortest connection between the image source and the receiver $\mathbf{x}_c$:

$$g = ((\mathbf{x}_{src} - \mathbf{x}_c) \cdot (\mathbf{x}_c - \mathbf{x}_{rec}))^{2.7} \quad (3)$$

In this study, only reflections of first order are simulated. Primary sources and reflected sources are simulated as omnidirectional sources.

Diffuse sources, e.g., background signals, or diffuse reverberation, are added in first order ambisonics (FOA) format. No distance law is applied to diffuse sound sources; instead, they have a rectangular spatial range box, i.e., they are only rendered if the receiver is within their range box, with a von-Hann ramp at the boundaries of the range box. Position and orientation of the range box can vary with time. The diffuse source signal is rotated by the difference between receiver orientation and box orientation.

In the acoustic simulation, receivers can be considered as virtual microphones, i.e., receivers return the output signals of the acoustic simulation. Two receiver types are used: A *spatial receiver* encodes the direction of the direct and reflected sound sources in 3rd order ambisonics, and adds the diffuse sound sources to the first order components of the receiver output[1]. *Omnidirectional receivers* with a single output are used to return the signal of the direct and reflected sources to the external diffuse reverberation generators. Both receiver types apply the distance and air-absorption mode to the source signals. The omnidirectional receivers can have a finite range box. If a source is within that range box of a receiver the distance law is only applied to the delay and not to the gain and air absorption model. This way it can be achieved to have all sources within a simulated room contributing with the same gain to the diffuse reverberation. Both receiver types can be restricted to render only sources within a range box, and they can also be restricted to render only direct point sources, mirrored point sources or diffuse sources.

In the proposed simulation framework, each room is simulated separately. Coupling is provided by 'sound portals' which are situated in the (door) openings connecting adjacent rooms. A sound portal has a virtual sound source attached radiating from the nearest point of the surface of the portal surface to the receiver. In the distance law, however, the distance between the sound portal and the source room center is added, to create a plausible fall-off rate. The sound radiated from the sound portal is gained from an omnidirectional receiver in the source room behind the portal surface.

Diffuse reverberation is rendered by an external tool. Each simulated room provides an omnidirectional (or optionally a directional) receiver. The output signal of that receiver is processed by a diffuse reverberation model [10] configured to match the specific room qualities. The first order ambisonics reverberation signal is then added as a diffuse input source.

3. EVALUATION METHODS

3.1. Simulated environment

The simulated environment was an office room ($4.43 \cdot 4.5 \cdot 3$ m^3) next to a long corridor ($30 \cdot 1.94 \cdot 2.5$ m^3). The sound source was always in the office room at a fixed position. Five static listening positions (see Figure 2) have been rendered: The first was close to the source in the office room, facing towards the sound source. The second position was closer to the door, but facing the same direction. The third listening position was exactly in the door between the office room and the corridor. The fourth and fifth position were in the corridor, with the source hidden by the wall. In the dynamic situations the position was interpolated linearly, within 20 seconds, resulting in a velocity between 0.14 and 0.35 m/s.

In the room simulation model, reflectors were placed at all walls, with the exception of the door. In the office, the direct sound

[1]The authors are aware of artifacts caused by playback of first order sources via higher order systems, caused by near field compensation order weights and coloration due to correlated sources played over many loudspeakers. However, both limitations do not play a major role in this setup, since first order sources are only used for diffuse and thus uncorrelated sounds, and near field compensation is not applied here.

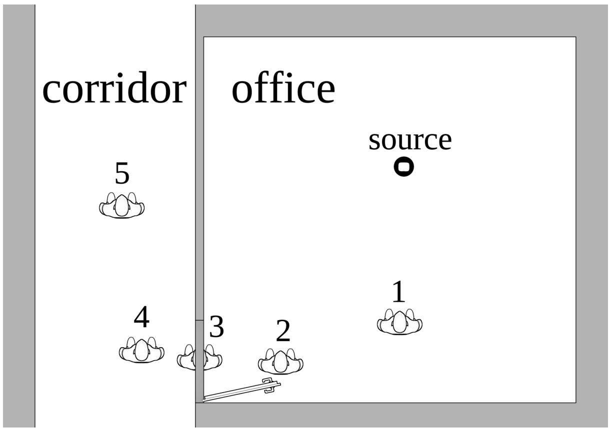

Figure 2: *Simulated environment, with listening positions marked by an artificial head.*

Position	1	2	3	4	5
Distance [m]	1.87	2.78	3.34	4.04	5.40
Azimuth [deg.]	-1.5	-32.4	-46.5	-97.3	-152.9

Table 1: Distance and direction between receiver and sound source. For positions 4 and 5, the distance is the sum of the distance receiver-sound portal and sound portal-source. The azimuth of positions 4 and 5 represents the direction to the sound portal.

position	1	2	3	4	5
simulated:					
L_l [dB]	-2.7	-10.2	-11.7	-11.3	(-15.3)
L_r [dB]	-1.9	-5.7	-1.4	9.6	(4.5)
$T_{60,1}$ [s]	0.16	0.19	0.13	0.02	0.09
$T_{60,2}$ [s]	0.42	0.43	0.48	0.59	0.60
ILD [dB]	-0.1	0.1	-1.1	0.4	-4.1
IACC	0.52	0.21	0.17	0.12	0.13
recorded:					
L_l [dB]	-4.2	-11.9	-11.3	-14.2	(-6.1)
L_r [dB]	-5.8	-5.0	-5.4	-3.7	(-12.8)
$T_{60,1}$ [s]	0.25	0.27	0.33	0.42	0.40
$T_{60,2}$ [s]	0.37	0.42	0.47	0.67	0.65
ILD [dB]	1.0	4.2	7.3	5.9	-1.9
IACC	0.32	0.08	0.13	0.08	0.09

Table 2: Physical characterization of the BRIR at the tested listening positions. L_l and L_r denote the 'liveness' [15] at the left and right ear, respectively. None of the positions 1 to 3 are within the critical distance. $T_{60,1}$ and $T_{60,2}$ are the early and late reverberation times [16]. The early reverberation time is lower for the simulated BRIRs than for the recorded BRIRs. Interaural level difference (ILD) suggest a small lateralization for the simulated method opposed to a large lateralization for the recorded BRIR. The interaural cross correlation (IACC) is slightly higher in the simulation. In the listening position 5 no direct sound was audible, thus the 'liveness' measures L_l and L_r have to be interpreted with care and are given in brackets.

and the early reflections were recorded by an omnidirectional receiver in the size of the office room. The output of this receiver was used as an input of the diffuse reverberation simulator. The output of the diffuse reverberation simulator was added to the listener when in the office room, and fed into the 'door' input, which was propagating it further into the corridor. The 'door' output was used as a directed sound by the listener receiver, and used as input of the diffuse reverberation simulator of the corridor. The von-Hann ramps of the receivers were 0.2 m long.

In Table 1 the distance and direction between receiver and source is given. For the positions 4 and 5, the sum of the distance receiver-sound portal and sound portal-source is given, as well as the direction between the receiver and the door.

3.2. Apparatus and test stimuli

All signals used in the subjective evaluation were pre-rendered. Stimuli were played back via a Sennheiser HD580 headphone. The levels could be adapted individually to match a comfortable level during the beginning of the measurement, and were kept fix during the measurement.

The receiver used for listening was a full periphonic 3rd-order ambisonics encoder. The receiver output was routed to an ambisonics decoder [11] to get the loudspeaker feeds. Binaural signals were generated by convolution with binaural head-related impulse responses (HRIR) of an artificial head for the 20 virtual speaker positions on the vertices of a regular dodecahedron. In the situation with static receiver (listener) positions binaural room impulse responses (BRIR) were generated by using a Dirac-delta pulse as the input signal of the primary source. Two anechoic test signals with a duration of 20 seconds, a speech signal (labeled 'speech') and a music recording [12] (labeled 'guit.'), were convolved with the BRIRs. As a reference condition, BRIRs have been recorded in the equivalent real room. Room impulse responses were measured using an omnidirectional loudspeaker. For recording, an artificial head MK2 by Cortex with a respective measurement amplifier Manikin MK1 was used. The excitation signal was a logarithmic sweep [13], with starting and ending frequencies of 50 Hz and 18 kHz, respectively. See [10] for details. The same measurement procedure and equipment was used to obtain the HRIR database for binaural auralization.

The test signals with a moving listener are directly rendered using the same anechoic source input signals. The time-varying listening position was linearly interpolated between the five discrete positions. Additionally, for the dynamic situation, a visual simulation of the scenario was rendered as a video using a 3d-graphics tool [14], with the same spatial properties of the simulated environment as in the acoustic simulation.

3.3. Physical characterization of the BRIR

To characterize the simulated and recorded BRIR, the 'liveness' L [15], i.e., the ratio between the direct sound to the reverberant sound, has been derived from the impulse responses. This is related to the critical distance, at which L equals 0, i.e., values of $L \geq 0$ represent a receiver within the critical distance. Additionally, the early and late reverberation time, $T_{60,1}$ and $T_{60,2}$, measured after [16], are given for comparison. Interaural level difference (ILD) and the interaural cross correlation (IACC, [17]) as used by [10] are also provided. The data are given in Table 2.

3.4. Spatial awareness

To evaluate the spatial quality of the simulated room coupling, the auditory spatial awareness [7] was measured using an identification task: In a graphical interface, the listener was able to switch between the stimuli for the five listening positions, without knowing the order. The order of the stimuli was randomized. They were time-aligned and repeated, i.e., by switching between the stimuli only the spatial configuration of the sound changed. The task was to assign the number of the listening position to each sound. The users were able to revise their decision at any time; when a listening position was assigned to all sounds the subjects were able to finally confirm their decision and end the measurement. A confusion matrix was calculated from the results. If the acoustic stimulus delivered a spatial awareness, then only the diagonal elements would differ from zero. As a measure of spatial quality, the accuracy, i.e., the sum of correctly identified stimuli divided by the sum of presented stimuli, was calculated. The individual accuracy for each stimulus and presentation method was calculated as well as the pooled accuracy for all subjects and test stimuli in each presentation method. Additionally to the accuracy, specific confusions bear the potential of revealing artifacts of the simulation method.

The spatial awareness measurement on a ranked path, as applied here, can be regarded as an indirect measure of the combination of source localization and distance perception.

Both, recorded BRIR and simulated BRIR, were tested.

3.5. Subjective rating

Absolute rating of naturalness was asked for during the playback of the simulated video in combination with the simulated sounds, for the speech signal and the music signal. The subjects were ask to rate the overall naturalness, the naturalness in the office room, in the corridor, and during the transition. The listeners knew the simulated position by watching the video representation.

3.6. Test subjects

Thirteen normal hearing listeners (average age 31 years, standard deviation 7 years) participated in the experiment.

4. RESULTS

Results of the effect of the simulation method on spatial awareness are shown in Fig. 3 and 4. Figure 3 shows the confusion matrix between presented and perceived listening position for simulated static room impulse responses (upper panels) and measured binaural room impulse responses (lower panels), for the two test stimuli.

All positions were detected correctly by more than 50% of the test subjects, for any reproduction method and test signal. However, with the simulation method there is a confusion noticeable between position 3 (door) and 4 (corridor, close to door). This is also represented by the decreased accuracy (median of the individual accuracy of 0.6 for simulated acoustics versus 1 for recorded BRIR, or pooled accuracy over all test signals and subjects of 0.68 for simulated acoustics versus 0.88 for the recorded BRIR). Figure 4 shows median values, quartile-ranges and extrema of the accuracy for the different test conditions. A two-way analyis of variance (ANOVA) revealed that the effect of the presentation method (simulation or BRIR) on accuracy is statistically significant (p=0.05); the test stimulus has no significant effect.

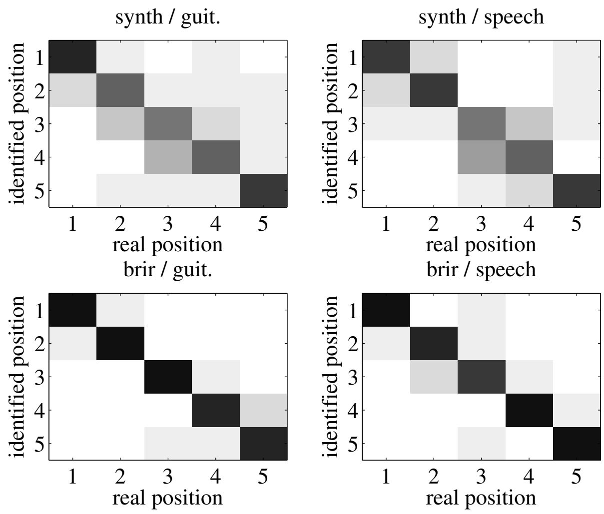

Figure 3: *Confusion matrices for the four tested conditions; results for simulated listening positions are shown in the upper panels, and for recorded BRIRs in the lower panels. For the simulated listening positions, a confusion between listening position 3 and 4 is noticeable.*

Absolute rating of naturalness (Figure 5) in the dynamic condition shows a moderate to high naturalness. The naturalness is perceived best in the office, i.e., when the direct sound of the source is audible. The 'corridor' listening position is rated slightly worse, which was also supported by some test subject comments mentioning that the sound level in the corridor was perceived to high compared to the sound level in the office. A two-way ANOVA with factors listening position and test stimulus [4x2] revealed a significant main effect of listening position (p=0.05). No significant effect of test stimulus was observed. Post-hoc comparison using Fisher's LSD criterion reveals that the rating in the office is significantly better than in all other tested conditions (p=0.05).

5. DISCUSSION

The evaluation of the simulation method for coupled rooms shows deficits when being compared to measured binaural room impulse responses. While extreme positions were correctly identified in most cases, it was more difficult to distinguish between the listening positions close to the door. This might be caused by the fact that the door simulation source radiates the sound of the office room with an insufficient falloff rate, and that the direct sound is not disappearing appropriately when moving from listening position 3 to position 4. It is primarily this confusion which decreases the accuracy to values around 0.6. The confusion between the first and the second listening positions might be a hint that the first early reflections are to strong, and that higher order reflections are needed for a better distinction between the listening positions within the room. This is also supported by the early reverberation time, which is smaller for the simulation than for recorded BRIR, and by the interaural cross correlation, which is higher for the simulation. Further more, the physically measured ILD does not reflect the lateralization of the source when moving from position 1 to position 3. Also a coupling from the corridor back to the

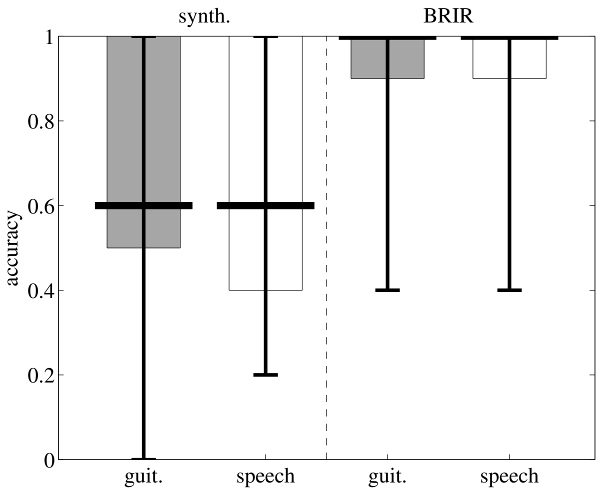

Figure 4: *The accuracy of the simulated (left half) and recorded (right half) listening positions. Median values (thick black lines), inter-quartile ranges (gray bars) and extrema across all test subjects are shown. An accuracy of 0.6 is reached if a single pair of listening positions is exchanged, i.e., most subjects were able to identify three listening positions correctly with the simulated listening positions, and identified all listening positions correctly with the recorded BRIR.*

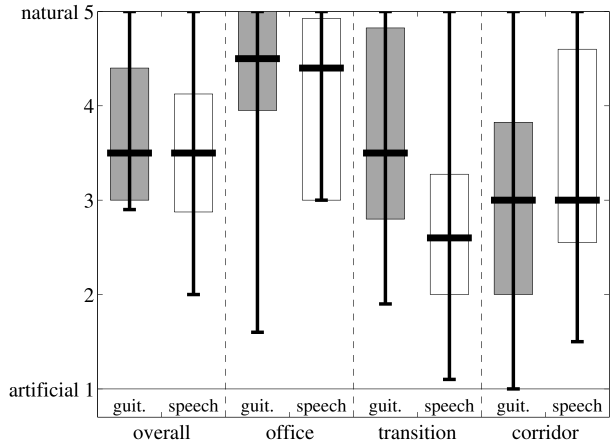

Figure 5: *Rating of the naturalness of the dynamic simulation. A video with a visual simulation of the environment from the listener's perspective was presented together with the simulated audio for either a music test signal ('guit.') or a speech test signal ('speech'). The subjects were ask to rate the overall naturalness and the naturalness of the simulation in the office, during the transition through the door, and in the corridor. Naturalness was rated on a five-point scale.*

office room, which was not applied here, may improve the spatial awareness.

In general the method of measuring the spatial quality in terms of spatial awareness is suitable to show differences between simulation methods, even if the stimuli differ in their spectra. A direct comparison is likely to be dominated by the spectral differences of the stimuli; this effect can be excluded using the spatial awareness test. On the other hand, spatial awareness is only one of many facets of spatial reproduction methods; thus this test method can not supersede other tests. Specifically, the presented method can only be applied to static positions, whereas the other test method of this study, absolute rating of naturalness, can also be applied to dynamic stimuli.

Some test subjects commented on the dynamic simulation after the experiment: Three subjects mentioned that the overall level of the test signal in the corridor was to high in relation to the direct sound in the office. High values of the 'liveness' measure for the simulation in the corridor may indicate that the virtual source in the door was too dominant. This level difference may have been the reason for the slightly lower naturalness rating in the corridor. Also some test subjects remarked that the transition between the rooms was difficult to rate for the speech signal, because there was a gap between sentences exactly in the door. This is likely to be responsible for the lower rating of the speech signal in the transition.

The current evaluation showed satisfying plausible reproduction for the given path. Future research will investigate whether the proposed subjective evaluation method and the BRIR synthesis method are more generally applicable to arbitrary paths in different rooms.

6. CONCLUSIONS

A computationally highly efficient real-time simulation method for transitions between coupled rooms was presented. The performance of the simulation method has been evaluated by measures of spatial awareness and by absolute subjective ratings. Although in absolute ratings the overall naturalness is rated high, the spatial awareness test reveals specific problems in the suggested simulation method. The presented data proof the function and applicability of the method and provide valuable input for further improvement of the simulation method. An improved version should better reproduce the ILD and the early reverberation time. The ILD could be improved by attenuating the diffuse sources in the direction of nearby walls. The early reverberation time can be improved by adding higher order reflections in the mirror source model. The unnaturally high amount of direct sound at the listening positions in the corridor can be improved by excluding the direct sound in the sound radiated by the sound portal in the door opening. A more systematic evaluation would require to investigate more paths, and different rooms.

7. ACKNOWLEDGMENTS

This study was funded by the DFG FOR 1732 "Individualisierte Hörakustik".

8. REFERENCES

[1] "Ease," http://ease.afmg.eu/.

[2] "odeon," http://www.odeon.dk/.

[3] Jont B Allen and David A Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, pp. 943, 1979.

[4] A. Krokstad, S Strøm, and S. Sørsdal, "Calculating the acoustical room impulse response by the use of a ray tracing technique," *J. Sound Vib.*, vol. 8, no. 1, pp. 118–125, 1968.

[5] Michael Vorländer, "Simulation of the transient and steadystate sound propagation in rooms using a new combined raytracing/image-source algorithm," *J. Acoust. Soc. Am.*, vol. 86, pp. 172–178, 1989.

[6] B.-I. Dalenbäck and M. Strömberg, "Real Time Walkthrough Auralization – The First Year," Tech. Rep., CATT (Dalenbäck), Valeo Graphics (Strömberg), 2010.

[7] Barry Blesser and Linda-Ruth Salter, *Spaces Speak, Are You Listening?*, The MIT Press, 2007.

[8] Giso Grimm, Graham Coleman, and Volker Hohmann, "Realistic spatially complex acoustic scenes for space-aware hearing aids and computational acoustic scene analysis," in *16. Jahrestagung der Deutschen Gesellschaft für Audiologie*, Rostock, Germany, 2013.

[9] Jyri Huopaniemi, Lauri Savioja, and Matti Karjalainen, "Modeling of reflections and air absorption in acoustical spaces a digital filter design approach," in *Applications of Signal Processing to Audio and Acoustics, 1997. 1997 IEEE ASSP Workshop on*. IEEE, 1997, pp. 4–pp.

[10] Torben Wendt, Steven van der Par, and Stephan D. Ewert, "Perceptual and room acoustical evaluation of a computational efficient binaural room impulse response simulation method," in *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*, 2014, submitted.

[11] Fons Adriaensen, "Ambdec 0.5.1 - first, seond and third order Ambisonics decoder," `http://kokkinizita.linuxaudio.org/linuxaudio/`, 2011.

[12] Michio Woirgardt, Philipp Stade, Jeffrey Amankwor, Benjamin Bernschütz, and Johannes Arend, "Cologne university of applied sciences – anechoic recordings," `http://www.audiogroup.web.fh-koeln.de/`, 2012.

[13] Angelo Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*, 2 2000.

[14] "blender," `http://www.blender.org/`, 2013, Stichting Blender Foundation, Amsterdam, the Netherlands.

[15] JP Maxfield and WJ Albersheim, "An acoustic constant of enclosed spaces correlatable with their apparent liveness," *The Journal of the Acoustical Society of America*, vol. 19, pp. 71–79, 1947.

[16] Manfred R Schroeder, "New method of measuring reverberation time," *The Journal of the Acoustical Society of America*, vol. 37, pp. 409–412, 1965.

[17] T. Hidaka, L. L. Beranek, and T. Okano, "Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls," *The Journal of the Acoustical Society of America*, vol. 98, pp. 988–1007, 1995.